

PUSASQF405
APPLICATIONS OF IT - BASICS OF PYTHON

Time: 2 hours
Total Marks: 60

Note:

- 1. The candidate has the option to either question 3A or question 3B. Rest all questions are mandatory.**
- 2. Numbers to the right indicate full marks.**
- 3. The candidates will be provided with the formula sheet and graphs (if required) for the examination.**
- 4. Use of approved scientific calculators is allowed.**

Q1. Attempt All questions.

15 Marks

A.

5 Marks

Given a 2D list representing survey responses (each row is a survey answer), count the number of occurrences of each answer (e.g., Y/N) and output the result in a dictionary.

Input: ['y', 'n', 'n', 'y', 'n', 'y', 'y', 'y', 'y', 'n']

Output: {"y": 6, "n": 4}

B.

5 Marks

Write a program which prompts the user to input principal, the rate is 7.5% annually, calculate the final amount (after 15 years). Plot the interest amount over 9 years on a bar chart. The formula is: $CI = P(1+R/100)^T$

C.

5 Marks

Given two dataframe, perform inner and outer joins using the merge function in pandas. Also, perform inner join after filling in the missing values

```
df1 = pd.DataFrame({ 'Month': [April, May, June,
July], 'Value': [100, 200, 300, 400] })
df2 = pd.DataFrame({ 'Month': [April, May, July],
'Amount': [50, 150, 250] })
```

Q2. Answer the following questions, based on the mall_customers.csv.

15 Marks

A.

5 Marks

Read the csv file and store them in df1 and write a program that filters and displays the customers with an annual income between 16k\$ and 19k\$, and show their CustomerID, Gender, Age, and Spending Score?

B.

5 Marks

Write a program that calculates the average annual income and average spending score for each gender. Output the result in a dictionary where the key is the gender and the values are the average income and spending score?

C.

5 Marks

Write a program that bins customers into three categories based on their annual income:

- (1) Low income ($\leq 15k\$$),
- (2) Medium income (16k\$ - 20k\$), and
- (3) High income ($> 20k\$$).

Provide no. of customers and avg spending score in each bin

Q3. Attempt question 3A or question 3B.

30 Marks

A.

30 Marks

Load the Credit Card dataset. The dataset contains information about the application of credit card related factors. Predict the acceptance/rejection with all other factors given in the dataset using Random Forest

a. Read the dataset and split the data in X and Y.

i) Identify independent variables consider it X (1)

ii) Identify dependent variables consider it Y (1)

b. Perform feature engineering, fill all missing values and Perform label encoding or One hot encoding whenever required. (6)

c. Perform train-test split with test ratio 30% (3)

d. Perform Random Forest on train dataset. (5)

e. Predict the values on Test dataset. (5)

f. Provide the value for Accuracy score and other classification matrix like precision, recall etc. (5)

g. Plot AUC ROC curve and explain how it is different from other classification metrics (4)

OR

B.

30 Marks

Load the car insurance premium dataset. The dataset contains information about the car insurance premium and the factors that affect it. Predict the Insurance Premium (\$) using Generalized Linear Model.

- a. Read the dataset and split the data in X and Y.
 - i. Identify independent variables consider it X (1)
 - ii. Identify dependent variables consider it Y (1)
- b. Perform feature engineering, fill all missing values and Perform label encoding or One hot encoding whenever required. (6)
- c. Perform feature importance using Lasso & ridge regression techniques, eliminate features which are less important (6)
- d. Create a train-test split with 10% test ratio. (3)
- e. Run the Generalized Linear Model. (5)
- f. Predict values for test dataset. (5)
- g. Provide the RMSE & R square value (3)